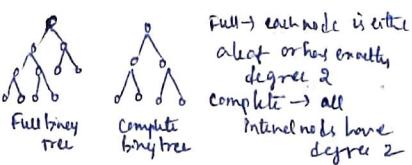


Tree :- degree \rightarrow no. of children
 depth \rightarrow length of simple path from root to x
 Level \rightarrow all nodes at the same depth
 height \rightarrow length of the longest simple path from x downwards to some leaf node.



- \rightarrow at most 2^d nodes at level $1 \leq d$ of binary tree
 - \rightarrow with depth d has at most $2^{d+1}-1$ nodes
 - \rightarrow with n nodes has depth at least $\lceil \lg n \rceil$
- $$n \leq \sum_{k=0}^{d-1} 2^k = \frac{2^d - 1}{2-1} = 2^{d+1}-1$$

Heap \rightarrow complete binary tree with structural property: all levels are full, except possibly the last one, which is filled from left to right.
 Other (heap) property: for any node x parent(x) $\geq x$.

Array Representation of Heaps =
 Root $A[i]$, Node i's $A[i]$, Left child of node $i = A[2i]$, Right child of node $i = A[2i+1]$
 Parent of node $i = A[\lfloor i/2 \rfloor]$
 elements in subarray $A[\lceil \lg n \rceil + 1 \dots n]$ are leaves
 max heaps (largest element at root)
 $\rightarrow A[\text{parent}(i)] \geq A[i]$
 min heaps (smallest element at root)
 $\rightarrow A[\text{parent}(i)] \leq A[i]$

Max Heaps (A[i])
 {
 $l = \text{left}(i); r = \text{right}(i);$
 If ($l \leq \text{heap_size}(A) \text{ and } A[l] > A[i]$)
 largest = l;
 else largest = i;
 If ($l \leq \text{heap_size}(A) \text{ and } A[l] > \text{largest}$)
 largest = l;
 If (largest != i)
 Swap (A[i], largest);
 Heapsify (A, largest);
 }
 Assumptions: Left & Right Siblings of i are max heaps. $A[i]$ may be smaller than its children.

Building a Heap \rightarrow convert an array $A[1 \dots n]$ into a max heap ($n = \text{length}(A)$)
 elements in the Subarray $A[\lceil \lg n \rceil + 1 \dots n]$ are leaves.
 Apply max Heapsify on elements between $1 \dots \lceil \lg n \rceil$

Build_max_Heap(A) $\Theta(\lg n)$
 $n = \text{length}(A)$
 for $i \leftarrow \lceil \lg n \rceil$ down to 1
 do MaxHeapsify (A, i, n)

Heapsify:
 Build a maxheap from the array
 Swap the root with the last element in the array
 Discard the last node by decreasing its heap size
 call MaxHeapsify on the new root
 Repeat this process until only one node remains.

Analyzing Heapsify:
 Build max Heapsify $\Theta(n)$
 for $i \leftarrow \text{length}(A)$ down to 2
 do exchange $A[i] \leftrightarrow A[\text{parent}(i)]$
 Max Heapsify ($A[1 \dots n-1]$) $\Theta(\lg n)$

$\Theta(n) + (n-1)\Theta(\lg n) = \Theta(n \lg n)$

Heap Maximum (A)
 return $A[1]$ $\Theta(1)$

Heap Extract Max (A[n])
 If $n < 1$
 Then error "heap underflow"
 max $\leftarrow A[1]$
 $A[1] \leftarrow A[n]$ $\Theta(\lg n)$
 Max Heapsify ($A[1 \dots n-1]$)
 return max

Exchange the root element with the last
 Decrease the size of the heap by 1 element
 Call max Heapsify on the new root, in a
 heap of size $n-1$

Heap Increase Key (A, i, key)
 If key $\leq A[i]$
 Then error "new key is smaller than current key"
 $A[i] \leftarrow \text{key}$
 while $i > 1 \text{ and } A[\text{parent}(i)] < A[i]$
 do exchange $A[i] \leftrightarrow A[\text{parent}(i)]$
 $i \leftarrow \text{parent}(i)$ $\Theta(\lg n)$

Increment the key of $A[i]$ to its new value
 If the max heap property does not hold anymore. Traverse up to the root to find the proper place of the newly increased key.

Max Heapsify (A, key, n)
 $\text{heap_size}[A] \leftarrow n+1$ $\Theta(\lg n)$
 $A[n+1] \leftarrow -\infty$

Heap Increase Key (A, n+1, key)

Expand the max heap with a new element whose key is $-\infty$
 Call Heapsify to set the key of the new node to its correct value and maintain the max heap property.

Quicksort (A[p..r]) $\frac{A[p..r]}{A[q+1..n]}$

Quicksort (A[p..r])
 If $p \geq r$
 Then $q \leftarrow \text{partition}(A[p..r])$
 Quicksort (A[p..q])
 Quicksort (A, q+1, r)

Partition (A, p, r)
 $x \leftarrow A[p]$
 $i \leftarrow p+1$
 $j \leftarrow i+1$
 while True
 do repeat $j \leftarrow j-1$ $A[p..i] \leftarrow A[p..i-1]$
 until $A[i] \leq x$
 do repeat $i \leftarrow i+1$
 until $A[i] \geq x$ $\Theta(n)$
 If $i < j$
 exchange $A[i] \leftrightarrow A[j]$

else return j

$T(n) = T(q) + T(n-q) + 1$

Quicksort with Care \rightarrow
 $T(n) = T(1) + T(n-1) + n \Rightarrow \Theta(n^2)$
 Best case = $q = n/2$
 $T(n) = 2T(n/2) + n \Rightarrow \Theta(n \lg n)$

Randomised partition (A, p, r)

i $\leftarrow \text{Random}(p, r)$
 exchange $A[i] \leftrightarrow A[r]$

return partition (A, p, r)

Randomised Quicksort (A, p, r)

If $p > r$
 then $q \leftarrow \text{Randomised partition}(A, p, r)$

Randomised Quicksort (A, p, q)

Randomised Quicksort (A, q+1, r)

Another partitioning: Lomuto's partition

Pivot $x = A[q]$
 $A[p \dots q-1]$ smaller than x
 $A[q+1 \dots r]$ is greater than x
 Pivot not included in any of the arrays.

Randomised Quicksort (A, p, r)

If $p > r$
 then $q \leftarrow \text{Randomised partition}(A, p, r)$
 Randomised Quicksort (A, p, q-1)
 Randomised Quicksort (A, q+1, r)

partition (A, p, r)

$x \leftarrow A[r]$

$i \leftarrow p-1$

for $j \leftarrow p$ to $r-1$

do If $A[j] \leq x$

then $i \leftarrow i+1$

exchange $A[i] \leftrightarrow A[j]$

exchange $A[i+1] \leftrightarrow A[r]$

return $i+1$

Total work done $\Theta(n^2) = \Theta(n \lg n)$

Expected value $E[X] = \sum x P(x = x)$

Indicator random value $I\{A\}$

$\mathbb{E}[I\{A\}] = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$

Expected value of an indicator random variable $X_A = \mathbb{E}[I\{A\}] \Rightarrow \mathbb{E}[X_A] = P\{A\}$

Total number of comparisons:

$\Rightarrow X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$ $\frac{n-1}{2} \dots n-1$

$\Rightarrow E[X] = E\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[X_{ij}]$ $\frac{n-1}{2} \dots n-1$

$= \frac{1}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} = \frac{1}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{n-i-1}{2}$

$\leq \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \left(\frac{2}{k}\right) (\lg n) \Rightarrow \Theta(n \lg n)$

All all comparison sets with $\frac{n-1}{2} \dots n-1$
 Lower bound Decision tree $\Omega(n \lg n)$

$n! \leq 2^n \leq \log(n!) \leq n$

Stirling's approximation tells $n! \geq \left(\frac{n}{e}\right)^n$

$\Rightarrow n! \geq \lg\left(\frac{n}{e}\right)^n \geq \lg n - \lg e = \Theta(n \lg n)$

Sorting in Linear time :-

Counting sort :- no comparisons b/w elements
→ For each element x , find the number of elements $\leq x$
→ place x into its correct position in the array.

→ $\Theta(n)$ allocation $\Theta(n)$ insert in B.

Counting sort (A, B, n, k):

for $i \leftarrow 0$ to n
do $c[i] \leftarrow 0$ } $\Theta(n)$
for $j \leftarrow 1$ to n
do $c[A[j]] \leftarrow c[A[j]] + 1$ } $\Theta(n)$
for $i \leftarrow 1$ to n
do $c[i] \leftarrow c[i] + c[i-1]$ } $\Theta(n)$
for $i \leftarrow n$ down to 1
do $B[c[A[i]]] \leftarrow A[i]$ } $\Theta(n)$
 $c[A[i]] \leftarrow c[A[i]] - 1$ } $\Theta(n)$

Radix sort :- represents keys as d -digit numbers in some base k

e.g. key = $x_1 \dots x_d$ where $0 \leq x_i \leq k-1$

key = 15

$key_1 = 15, d = 2, k = 10$ where $0 \leq x_1 \leq 9$

$key_2 = 1111, d = 4, k = 2$ where $0 \leq x_1 \leq 1$

Assumptions :- $d = O(1)$ & $k = O(n)$

→ For a d digit number, Sort the least significant digit first using the stable sort algorithm.

→ continue sorting on the next least significant digit, until all digits sorted

→ running time $O(d(n+k))$

order statistics :-

i th order statistic c in a set of n elements is the i th smallest element

→ minimum is 1st order statistic

→ max is n th order statistic

→ median is $n/2$ th order statistic c .

→ never Theta $\Theta(n)$ medians.

We can find the min & max with less than twice of cost \rightarrow yes

walk through elements by pairs

→ compare each element in pair to other

→ compare the largest to max & smallest to min

→ Total cost : 3 comparisons per 2 elements

Randomised selection :-

key idea: Use Partition from quicksort

But only need to examine one subarray

Saves running time $O(n)$

$q = \text{Randomized partition}(A, p, r)$

$\boxed{\begin{array}{|c|c|c|} \hline & \text{C}[A[1]] & \dots & \text{C}[A[q]] \\ \hline & p & & q \\ \hline \end{array}}$

Randomized select (A, p, r, i)

If $i = r$ then return $A[r]$;

$q = \text{Randomized partition}(A, p, r)$

$k = q - p + 1$;

If $i = k$ then return $A[q]$;

If $i < k$ then

return Randomized select ($A, p, q-1, i$);

else

return Randomized select ($A, q+1, r, i - k$);

Analyzing Randomised Selection :-

Worst case : partition always $O(n)$

$T(n) = T(n-1) + O(n) = O(n^2)$

Best case : suppose $q = 1$ partition

$T(n) = T(n-1) + O(n) = O(n)$

$= O(n)$

Average case :-

$$\begin{aligned} T(n) &\leq \frac{1}{n} \sum_{k=0}^{n-1} T(\max(k, n-k-1)) + O(n) \\ &\leq \frac{2}{n} \sum_{k=1}^{n-1} T(k) + O(n) \\ &\leq \frac{2}{n} \sum_{k=1}^{n-1} ck + O(n) \\ &= \frac{2c}{n} \left(\sum_{k=1}^{n-1} k - \sum_{k=1}^{n-1} \frac{k}{2} \right) + O(n) \\ &= \frac{2c}{n} \left(\frac{1}{2}(n-1)n - \frac{1}{2} \left(\frac{n-1}{2} \right) \frac{n}{2} \right) + O(n) \\ &= c(n-1) - \frac{c}{2} \left(\frac{n-1}{2} \right) + O(n) \\ T(n) &\leq cn - c - \frac{cn}{4} + \frac{c}{2} + O(n) \\ &= cn - \frac{cn}{4} - \frac{c}{2} + O(n) \\ &\leq cn \end{aligned}$$

Search problem :- search x such that $x = A[i]$

Direct addressing :- Assumptions

key values are distinct

Each key is drawn from universe $U = \{1, \dots, m\}$

ideal :- store the items in array indexed by keys.

operations :-

Direct address search (T, k)

return $T[k]$ $O(1)$

Direct address insert (T, x)

$T[key[x]] \leftarrow x$ $O(1)$

Direct address Delete (T, x)

$T[key[x]] \leftarrow \text{NIL}$ $O(1)$

hash table :-

use function h to compute the slot for each

store element in slot $h(k)$

$h: U \rightarrow \{0, 1, \dots, m-1\}$

collisions :- For a given set of k keys

$f(k) \leq m$, collisions may not

happen, depending on the hash function

$f(k) > m$ collisions will definitely happen

chaining :-

put all elements that hash to the same

slot into a linked list

\rightarrow slot 3 contains a pointer to the head

of the list of all elements that hash to

chained hash insert (T, x)

$O(1)$ insert x at the head of list $T[h(key)]$

chained hash delete (T, x)

delete x from the list $T[h(key)]$

chained hash search (T, k)

search for element with key k in list

$T[h(key)]$

running time \propto length of the list of elements

in slot $h(k)$

Analysis of hashing with chaining

worst case :- All n keys hash to same slot

$\rightarrow O(n)$ + time to compute hash function

Average case :- Depends on hash function

Simple uniform hashing assumption

→ Any element is equally likely to hash

into any of m slots

Probability of collision $P(h(x)) = h(n) \frac{1}{m}$

Length of list $T[i] = n_i$, $i = 0, \dots, m-1$

num of keys in slot $n = n_0 + \dots + n_{m-1}$

Avg value of $n_j = E[n_j] = d = n/m$

load factor of hashtable $d = n/m$

n = number of elements stored in table

m = no of slots in table = no of linked list

d can be $\leq, =, \geq 1$

case 1 unsuccessful search item not stored in table takes expected time $O(1+d)$ \propto average length of chain

case 2 successful search $O(1+d)$ \propto length of chain

insert to the hashtable, \propto length of chain

$E \left(\frac{1}{n} \sum_{i=1}^n (1 + \sum_{j=1}^n x_{ij}) \right) = \frac{1}{n} \sum_{i=1}^n (1 + E[x_{ij}])$

$= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{1}{m} \sum_{j=1}^m 1 \right) = 1 + \frac{1}{nm} \sum_{i=1}^n (n-1)$

$= 1 + \frac{1}{nm} (n^2 - \frac{n}{2}) = 1 + \frac{n-1}{2n}$

Total time $= O(2 + \alpha_1 + \alpha_2 n) = \frac{1 + \alpha_1 + \alpha_2}{2n}$

Division method (hash function)

map a key k to one of the m slots by taking the remainder of k divided by m

$h(k) = k \bmod m$

multiplication method

multiply key k by constant α where α is a whole no

extract the fractional part of αk

multiply the fractional part by m

Take the floor of the result

$h(k) = \lfloor m(\alpha k - \lfloor m\alpha k \rfloor) \rfloor = \lfloor m(\alpha k \bmod 1) \rfloor$

universal hashing :-

Select a hash function at random, from a designed class of functions at the beginning of execution.

Designing a universal class of hash functions

choose a prime number p large enough

so that every possible key k is in range

$[0, \dots, p-1]$ $2p = [0, \dots, p-1]$

$\exists h_k(k) \in [(ak+b) \bmod p] \text{ mod } m$

$\forall a \in \mathbb{Z}^*$ and $b \in \mathbb{Z}$

The family of all such hash functions is said

$H_p, m = \{h_{a,b} : a \in \mathbb{Z}^* \text{ and } b \in \mathbb{Z}\}$

Binary search tree :-

If y is subtree of x then $key[y] \leq key[x]$

If y is right subtree of x then $key[y] \geq key[x]$

$key[y] \leq key[x] \leq key[z]$

$key[y] \leq key[z] \leq key[x]$

preorder : root left right $5 \ 2 \ 3 \ 7 \ 9$

postorder : left right root $2 \ 5 \ 3 \ 9 \ 7 \ 5$

inorder : tree-walk (x)

$\& x \neq \text{NIL}$ $O(n)$

2nd order tree-walk ($left[x]$) $O(n)$

Print key (x)

2nd order tree-walk ($right[x]$)

Tree Search (x, k)

$\& x = \text{NIL}$ $k = \text{key}[x]$

then return x

$\& k < \text{key}[x]$

then return Tree-Search ($left[x]$, k)

else return Tree-Search ($right[x]$, k)

Tree minimum (x)

while $left[x] \neq \text{NIL}$ $O(n)$

do $x \leftarrow left[x]$

return x

Tree maximum (x)

while $right[x] \neq \text{NIL}$ $O(n)$

do $x \leftarrow right[x]$

return x

Successor - Successor (x) = y , such that
key [y] is the smallest key $>$ key [x])
Tree - Successor (x):
if right [x] ≠ NIL
then return Tree-minimum(right [x])
 $y \leftarrow p[x]$
while $y \neq \text{NIL}$ & $x = \text{right}[y]$
do $y \leftarrow y$ $O(n)$
 $y \leftarrow p[y]$
return y

Predecessor - predecessor (x) = y such that
key [y] is the biggest key $<$ key [x])

Tree Insertion = Idea: If key [x] < v
move to the right child x
else move to the left
else if NIL
do $y \leftarrow x$
if key [y] < key [x]
then $x \leftarrow \text{left}[x]$
else $x \leftarrow \text{right}[x]$
 $y \leftarrow \text{NIL}$
 $x \leftarrow \text{root}[T]$
while $x \neq \text{NIL}$
do $y \leftarrow x$
if key [y] < key [x]
then $x \leftarrow \text{left}[x]$
else $x \leftarrow \text{right}[x]$
 $p[x] \leftarrow y$
if $y = \text{NIL}$
then $\text{root}[x] \leftarrow y$
else if key [y] < key [x]
then left [y] < 2
else right [y] < 2
point x to the downgrad path
(current node)
point y : parent of ("bailing point")

Deletion:
case 1: z has no children

Delete z by making the parent of z point
to NIL

case 2: z has one child

Delete z by making the parent of z point
to z 's child, instead of z

case 3: z has 2 children

$\rightarrow z$'s successor (y) is minimum node in z 's
right subtree
 $+ y$ has either no children or one right child
(but no left child)

\rightarrow Delete y from the tree (via case 1 or case 2)

\rightarrow Replace z 's key and satellite data with
 y 's

Insertion \rightarrow stable $\rightarrow O(n^2)$ worst & avg, $O(n^2)$ stable

Bubble \rightarrow stable $\rightarrow O(n^2)$

Radix \rightarrow stable $\rightarrow O(nk)$

Counting \rightarrow stable $\rightarrow O(n+k)$

Merge \rightarrow not stable, stable $\rightarrow O(nlg n)$

Heap \rightarrow unstable $\rightarrow O(nlg n)$

Quick \rightarrow unstable $\rightarrow O(n^2)$ worst, $O(nlg n)$ best

Selection \rightarrow unstable $\rightarrow O(n^2)$

Loop invariant of insertion:
At start of each iteration of the for loop of lines 1-8 the subarray $A[1 \dots j-1]$ consists of the elements originally in $A[1 \dots j-1]$ but in sorted order
Initialization: Just before first iteration $j=2$: The subarray $A[1 \dots j-1] = A[0:j]$ is sorted. Maintenance: while in inner loop moves $A[i-1], A[i], A[i-1]$ and so on by one position to the right until the proper position for key is found
- At that point the value of key is placed into this position
Termination: The outer loop ends when $j = n+1 \Rightarrow j-1 = n$
- Replace A with $j-1$ in the loop invariant \rightarrow The subarray $A[1 \dots n]$ consists of the elements originally in $A[1 \dots n]$ in sorted order.
Bubble sort: At the start of each iteration of the for loop of lines 1-4, the subarray $A[1 \dots i-1]$ consists of the $i-1$ smallest elements in $A[1 \dots n]$ in sorted order. $A[i \dots n]$ consists of $n-i+1$ remaining elements in $A[1 \dots n]$
Initialization: Initially the subarray $A[1 \dots i-1]$ is empty and trivially the i smallest element of the subarray
Maintenance: The execution of inner loop $A[i:j]$ will be the smallest element of subarray $A[i \dots n]$ and in the beginning of the outer loop $A[1 \dots i-1]$ consists of elements that are smaller than the elements of $A[i \dots n]$ in sorted order. So after execution of inner loop subarray $A[i \dots j]$ will consist of elements that are smaller than the elements of $A[i \dots n]$ in sorted order.
Termination: The loop terminates when $i = A.length$. At this point $A[1 \dots n]$ will contain all elements in sorted order.

Heap sort: At the start of each iteration of the for loop of lines 2-5 the subarray $A[1 \dots i]$ is a max heap containing the i smallest elements of $A[1 \dots n]$ and the subarray $A[i+1 \dots n]$ contains the $n-i$ largest elements of $A[1 \dots n]$ sorted.

$i=1$: The subarray $A[1 \dots n]$ is empty thus the invariant holds
 $i=2$: $A[1]$ is the largest element in $A[1 \dots i]$ and it is smaller than the element in $A[1 \dots n]$. When we put it in the pk position, the $A[1 \dots n]$ contains the largest elements sorted. Decreasing the heap size and calling max-heap turns $A[1 \dots n-1]$ into a max-heap. Decreasing i sets up the invariant for the next iteration.

$i=3$: After the loop $i=1$ this means that $A[1 \dots n]$ is sorted and $A[1]$ is the smallest element in the array which makes array sorted.

Radix sort: At the beginning of the for loop the array is sorted on the last $i-1$ digits

$i=1$: The array is trivially sorted on the last 0 digits
 $i=2$: Let's assume that the array is sorted on the last $i-1$ digits. After we sort on the i th digit the array will be sorted on the last i digits. It is obvious that elements with different digits in the i th position are ordered accordingly. In the ~~case~~ case of same digit, we still get a correct order, because we're using a stable sort and the element were already stored on the last $i-1$ digits.

$i=d$: the loop terminates when $i=d+1$ since the invariant holds, we have the numbers sorted on d digits.